WingWill Google Cloud

全方位 AI 應用安全防護指南

Google Cloud Model Armor：全方位 AI 安全防護指南 — 在執行階段為您的生成式與代理式 AI 提供頂級防護

● Model Armor 是一項專為提升 AI 應用程式安全性而設計的 Google Cloud 服務。它採用混合式縱深防禦機制，做為強大的 AI 防火牆，無論您是在 Google Cloud 或其他環境部署 AI，都能無縫保護包含 Gemini、Vertex AI、OpenAI、Anthropic 等多種大型語言模型 (LLM)。

核心特色與防護能力

抵禦惡意操弄與攻擊

主動偵測並阻擋提示詞注入 (Prompt Injection) 和越獄 (Jailbreak) 手法，防止惡意人士規避安全通訊協定，誘騙 AI 執行非預期動作或揭露機密。

精細的內容安全控管

提供自訂信賴度門檻（高、中、低），幫助企業精準過濾提示詞與回覆中的有害、不道德或不當內容（例如仇恨言論、騷擾、煽情露骨素材與危險主題），完美契合企業的風險承受能力。

頂級的敏感資料保護 (DLP)

無縫整合 Google Cloud 的 Sensitive Data Protection 服務。可自動偵測、遮蓋或阻擋提示詞和模型回覆中的個人識別資訊 (PII)、財務資訊、憑證及自訂私密資料，大幅降低資料外洩風險。

惡意軟體與網址攔截

不僅掃描文字，還能偵測 AI 提示詞、回覆以及上傳文件（支援 PDF、Word、Excel、CSV 等格式）中是否含有惡意網址與惡意檔案，避免網路釣魚與內嵌式威脅危害下游系統。

跨雲端與靈活部署

提供 REST API 介面，並支援無程式碼的內嵌防護機制，能輕鬆與 Vertex AI 網路服務擴充功能及 Apigee API 閘道等基礎架構整合，讓開發人員彈性運用於任何雲端環境。

立即聯繫羽昇國際，提升企業的雲端安全韌性！

企業常見應用場景

安全的客服聊天機器人

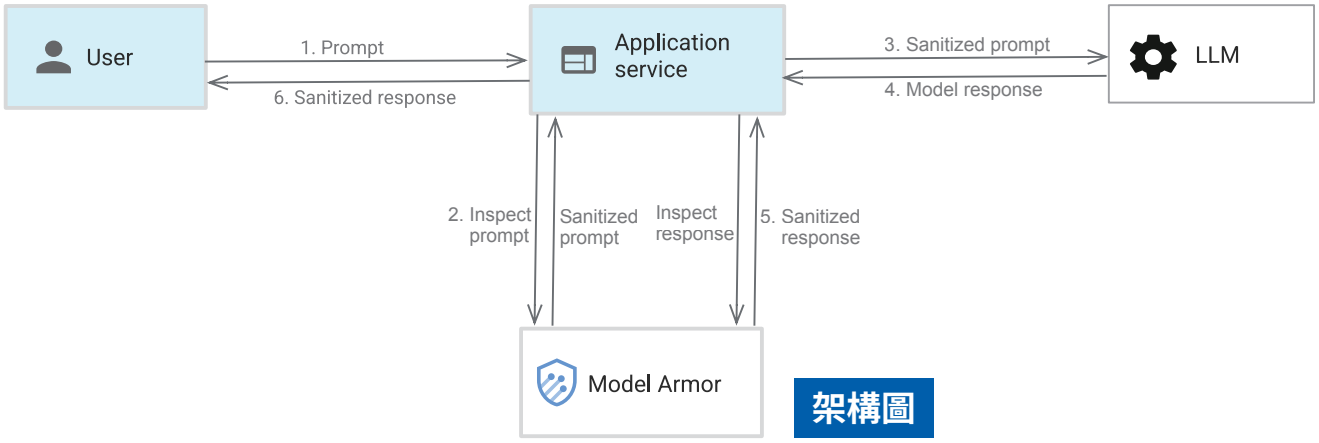
防止 LLM 輔助聊天機器人不慎洩露客戶敏感個資，或生成令人反感的回覆，全面守護品牌信譽。

企業內部知識 AI 代理

員工透過 AI 存取內部知識庫時，Model Armor 會套用嚴格的政策，確保機密資訊不會遭到意外揭露。

生成品牌安全內容

協助行銷團隊在草擬社群貼文或廣告文案時，確保生成的內容安全無虞、符合品牌形象，避免引發大眾負面情緒。



架構圖

彈性計價模式 (Pricing Model)

Model Armor 採用公開透明且極具彈性的計價架構，主要依據處理的詞元 (Token) 總數計費（包含 AI 提示與回覆中的分詞總數）。

具多種方案以滿足不同規模的企業需求：

即付即用方案 (Pay-as-you-go)

免綁約，超過免費額度後，完全依照實際使用的詞元數量計費。

多元整合選項

您可以選擇購買獨立的 Model Armor，或將其與 Security Command Center (SCC) Premium / Enterprise 版本整合。此外，這項防護功能也直接包含在 Gemini Enterprise 訂閱方案中，無需額外付費即可享有頂級防護。

企業訂閱方案

針對高用量需求，提供包含龐大月度詞元配額的訂閱涵蓋範圍。

羽昇國際 WingWill Int'l Co. Ltd.
(02) 26565630
service@wingwill.com.tw
企業雲端領航員，一站式整合服務 | Google Cloud 菁英合作夥伴

羽昇國際為零壹科技子公司，為雲端整合專家，專注於雲端應用市場，致力為各產業推動 IT 現代化，規劃最適切的雲端戰略，提供海內外客戶進行跨國多雲、混合雲、雲地整合以及資訊安全管理服務。

聯絡我們